148	retired epidemiologist	USA	6.1 The risk assessment process	6.1 The risk assessment process. 1) This is nearly the first place in this document where the Panel suggests that epidemiologic studies would be of any use in evaluating the relationship between pesticides and human health outcomes. It is almost like it was written by a different group. This section is only 4 sentences long. 2) Many epidemiologic studies of pesticides have information on exposure-response. To assume that this is useless for dose-response is incorrect. Epidemiologic studies of other exposures have long been used to provide information regarding dose-response and those of pesticides can also. It is usually not to the detail that is available in experimental studies of animals, but then epidemiology studies are on humans. The two disciplines provide important and complementary information. 3) It is true that it is hard to determine/estimate exposures in real-world situations. There is a sizable literature in epidemiology where this has been done form many substances, including pesticides. If "real-world" human exposures are so difficult to characterize, how is it possible to relate doses delivered to experimental animals to humans in any way. If this is not possible, it would seem that the experimental work is useless. *EFSA Response:* *Same text as comments #146 and 147.* 4) Why is the relevance of meta-analyses for risk assessment still limited? It would seem statistical procedures that combine data from individual studies would be preferable to simply looking at the results from individual studies? In fact, later in the document the Panel recommends the use of meta-analyses and pooling studies. *EFSA Response:* *Sentence in 1923: "their application [of meta-analyses] for risk assessment modelling is still limited" means that so far meta-analyses have not been systematically used for pesticide risk assessment. This Scientific Opinion intends to foster this use.* 5) I know this is standard stance for risk assessment, but it really makes no sense. Any animal study performed according to these standards is given a higher reliability than any study in the species of interest regardless of how well the epidemiology study is performed, or how much the animal study deviates from the human condition. How about studies in frogs or fish, are they still superior? *EFSA Response:* *The information on the toxic effects of pesticides for regulatory risk assessment is based on the results of a full set of experiments as required by Regulation (EC) 283/2013 conducted according to OECD guidelines Lines 2016-2018 under section 7.1).* 6) It is true that potentially confounding factors in animal studies can be controlled by study design when extrapolating results to humans? How is it possible to control for confounding from using the "wrong" species and often unrealistic exposures?

				EFSA Response: *The commenter is mixing two different things: the conduct of animal experimental studies (where all factors are controlled by study design) and their further extrapolation to humans.* 7) It is simply incorrect that epidemiologic studies cannot identify which exposures out of a mix of exposures are responsible for the outcome. This would be true if each individual in a study or all human populations have exactly the same mix of exposure. This does not occur. History makes it abundantly clear that epidemiology has able to evaluate a large number of specific exposures that are now established human health hazards and ALL humans have multiple exposures. *EFSA Response:* *The sentence in lines 2063-2066 states: "... in epidemiological studies there are always a variety of factors that may affect the disease outcome and confound the results. For example, when epidemiological data suggest that exposures to pesticide formulations are harmful they usually cannot identify what component may be responsible due to the complexity of accurately assessing human exposures to pesticides". In the light of many of epidemiological studies carried out so far, there is no clear and unambiguous information on what pesticide active substance out of the 480 authorised for use in the EU was actually assessed. Many times the broad term "exposure to pesticides" is used, or biomonitoring data on urine assessing levels of dialkylphosphate metabolites are not a clear indication of what organophosphate in particular was the population exposed to.* 8) The Panel indicates that studies that are capable of evaluating co-occurrence of multiple adverse outcomes, which happens in epidemiologic studies, are of value. Yet animal studies, which usually do not do this, are still superior to all epidemiology studies. *EFSA Response:* *The sentence in lines 2073-2075 states: "Risk assessment would benefit by developing procedures for evaluating evidence for co-occurrence of multiple adverse outcomes (Nachman et al., 2011), which is more in line with what happens in human setting". No indication for the superiority of animal studies is made in this sentence.* 9) Unmeasured confounders again. Please provide documentation of situations where findings/conclusions from epidemiology studies were later found to be entirely due to some uncontrolled confounding. The Panel is making a universal charge about confounding in epidemiology without providing any data. *EFSA Response:* *See comment #120.* *See comments #12 and 20 (the "wrong" species).*
149	German Federal Institute for Risk Assessment (BfR)	DEU	6.2 Assessment of the reliability of	Line 1764, page 40-41, Table 2: Study design and conduct: It is generally accepted that prospective cohort studies are of higher quality if high quality standards are followed. However, in cases where cohort studies are not appropriate (e.g. cases of rare diseases), a case control study would be the method of choice. In those cases, a case control study should be

			individual epidemiological studies	judged as "high". Exposure assessment: We could agree that use of validated questionnaires is a prerequisite for the "high" category. However, in some occasions matched controls could be preferred over random controls. In these cases the "high" category may also be appropriate. In addition, some studies require answers to chemical-specific exposure by proxies rather than by exposed subjects (e.g. exposure of children, subjects suffering from dementia). Those cases should also not lead to downgrading categorization from "high" to "moderate". Confounder control: It must be defined what is meant by "good control" of confounders. How should one discriminate between "good control" and "partial control"? This is not simple at all and must be clarified by examples. Reporting: Ideally, some identification of the substance/product (e.g. CAS number if applicable, purity if applicable, components of products etc.) would be useful. The approach was described as a "simple method …for evaluating and ranking human and experimental studies …" (line 201, p5). However, finding the most appropriate decision for the categories "high", "moderate" and "low" is not simple. A more comprehensive checklist would be helpful and may allow for better categorization. Overall: The examples above demonstrate that a more detailed assessment (e.g. checklist) could be an easier approach compared to only three categories "high" – "moderate" – "low". *EFSA Response:* *See also comment #150. The quality of epidemiologic research needs to be considered when evaluating epidemiology studies for pesticide risk assessment. Individual study evaluation is proposed to be performed by using the various considerations shown in Table 2 and the associated weight for each of the indicated parameters (low, medium, high). However, these considerations are generic and should not be regarded as a checklist. Specific considerations will be taken into account on a case by case basis, as would happen with the examples utilized by the commenter.* *The criteria included in Table 2 are based on those reported by the US-EPA (Framework for Incorporating Human Epidemiologic & Incident Data in Risk Assessments for Pesticides, Office of Pesticides Programs, December 28, 2016).* *As EFSA gains experience with integrating epidemiology studies into pesticide risk assessment, relevant adjustments to its evaluation approach will be made. Future revisions of this Opinion may consider the appropriateness of moving to a more detailed assessment.* *Good control of confounder means that potential confounding factors were appropriately identified and controlled for during the design phase of the study (e.g., by matching or restriction) or measured during the study conduct for further adjustment during the analysis phase (e.g., by stratification or multivariate analysis). Individual assessment of each study will ascertain whether potential confounding factors were carefully considered and the study adequately described how this was done.* *Text in lines 1747-1748 has been amended as following: "were potential confounding factors appropriately*

				identified and considered? How were they controlled for? Were the methods used to document these factors valid, reliable and adequate?".
150	ECPA	BEL	6.2 Assessment of the reliability of individual epidemiological studies	It is unclear why this section of the scientific opinion differs from Section 4.1, as both deal with study quality. There are many ways in which epidemiology study quality assessment should factor into EFSA's evaluations and regulatory decision making. For example, there should be minimum study quality requirements for studies selected for deriving quantitative limits (e.g. pesticide residue limits) and for studies selected for use in qualitative hazard and risk assessment and management (e.g. studies regarded as sufficient for consideration in pesticide re-registration decision making). Appendix A to our comments provides a set of guidelines; however, if these are used, additional guidance should be provided regarding how to fulfill each of the individual requirements. *EFSA Response:* *Section 4.1 deals with "Assessing and reporting the quality of epidemiological studies" under the context of how future epidemiological studies could be refined or improved for pesticide risk assessment. However, section 6.2 is intended to report the methodology for the assessment of the reliability of individual epidemiological studies. The Scientific Opinion intended to meet faithfully its terms of reference and also to be consistent by using the same systematic in the different chapters.* *Appendix A provided by the commenter describes a number of requirements that will be considered for improving the quality of reporting, similarly to the STROBE instrument.* Line 1752-1753 and table 2: It would be useful if the reporting requirement reflected the principle of transparency. STROBE adherence cannot be evaluated without a transparent research process. In addition to study quality, risk assessors must be able to ensure that the original objective of the studies being considered were similar to the one reported in the scientific literature. As mentioned in our comments above, the requirement of pre-registration and availability of the original study methodology should be promoted and encouraged. *EFSA Response:* *The principle of transparency has been considered as a reporting requirement.* *Text in lines 1752-1753 has been amended as following: "Is the reporting of the study adequate and following the principle of transparency and the guidelines of the STROBE statement (or similar tools)?"* Line 1763, table 2: In section 7.2 and 7.3 the PPR Panel describes principles for weighting of human observational and laboratory animal experimental data. However, there is no explicit guidance provided on the practical execution of weighting different sources of evidence. The study quality considerations in Table 2 and Figure 5 on weighting epidemiology observational studies are too imprecise and leave room for subjective interpretation (e.g. "case-control studies not adequately covering exposure or outcome assessment" is mentioned in column "moderate" as well as "low"; "adequate assessment of exposure, preferentially biomarker concentrations at individual level" is considered high in quality. However, if a biomarker study was measured using invalidated or non-replicable methods, exposure assessment should not be viewed as high quality. Therefore the inclusion of more detailed descriptions in table 2 would be helpful.

				EFSA Response: *This is Scientific Opinion, so a number of recommendations are given herein, but this is not intended to be a guidance.* *Figure 5 has been modified based on this and other comments.* *Table 2, first row (Study design and conduct) third column (Moderate) has been changed text to read: "Case-control studies. Prospective studies not adequately covering exposure or outcome assessment". Table 2, row 3 (exposure assessment), column 2 (high) has been changed to read: "Accurate and precise quantitative exposure assessment (human biomonitoring or external exposure) using validated methods". "Adequate assessment of exposure, preferentially biomarker concentrations at individual level" has been deleted.* Line 1770-1780: The PPR Panel provides three basic categories as a first tier to organize human data with respect to risk of bias and reliability - a) low risk of bias/high reliability, b) medium risk of bias/medium reliability, c) high risk of bias/low reliability. The rating is made according to how many of the quality factors provided in Table 2 ("many" or "most") have been met. However, in order to serve as practical guidance it should be more clearly defined what is specifically meant by "many" or "most". *EFSA Response:* *As mentioned above, this Scientific Opinion is not intended to be a guideline.* *Also, the response to comment #149 claimed that the considerations given in Table 2 are generic and should not be regarded as a checklist. Specific considerations will be taken into account on a case by case basis.* *As EFSA gains experience with integrating epidemiology studies into pesticide risk assessment, relevant adjustments to its evaluation approach will be made. Future revisions of this Opinion may consider the appropriateness of moving to a more detailed assessment.*
151	Ministero della Salute	ITA	6.2 Assessment of the reliability of individual epidemiological studies	Page 41 Lines 1770-1776: it is not clear which are the criteria for defining the 3 proposed categories (Expert judgement?). It is not even clear how Fig. 1 (at page 44) could be explanatory of these 3 categories. Is one additional figure missing? *EFSA Response:* *No specific criteria were proposed for defining how an individual epidemiological study can be classified in each of the three categories proposed for the study evaluation process. It depends on professional/expert. As indicated in comment #149, as EFSA gains experience with integrating epidemiology studies into pesticide risk assessment, relevant adjustments for a more detailed assessment will be considered in future revisions of this Opinion.* *Reference to Figure 1 in line 1776 has been deleted.*
152	Board for the authorization of Plant Protection Products and Biocide	NLD	6.2 Assessment of the reliability of individual epidemiolo	Line 1770 to 1780 discuss the three basic categories proposed to assess reliability of epidemiological studies. Reference is made to Figure 1 but this figure does not appear to be related to this section as it describes a fictitious example of meta analysis of two studies. Please check if the reference to figure 1 in this section is correct. *EFSA Response:* *Thank you for having identified this mistake. The reference to Figure 1 is not correct and has been removed from*